DECLARATION OF BENJAMIN KORNACKI ISO GOOGLE LLC'S MOTION FOR RELIEF REGARDING PRESERVATION

Redacted Version of Document Sought to be Sealed

1   **UNITED STATES DISTRICT COURT**

2   **NORTHERN DISTRICT OF CALIFORNIA, OAKLAND DIVISION**

3

4   CHASOM BROWN, et al., on behalf of
     themselves and all others similarly situated,

5   Plaintiffs,   Case No. 4:20-cv-03664-YGR-SVK

6   vs.

7   GOOGLE LLC,

8   Defendant.

9   PATRICK CALHOUN, et al.on behalf of
     themselves and all others similarly situated,

10

11   Plaintiffs,   Case No. 4:20-cv-05146-YGR-SVK

12   vs.

13   GOOGLE LLC,

14   Defendant.

15

16   **DECLARATION OF BENJAMIN KORNACKI**

17   1.   I am a Software Engineer employed by Google LLC. I have been employed at Google

18   since January 2013, and since then have been a member of the Ad Manager Reporting & Insights

19   Team. In my capacity as a member of the Ad Manager Reporting & Insights Team, I am responsible

20   for building pipelines to digest historical data and produce analytics and insights to our customers.

21   As part of my duties, I am familiar with the tools available at Google to access, search, and store

22   data in Google's ▮▮▮▮ Ads logs. I make this declaration based on personal knowledge and

23   information provided to me by Google colleagues, and if called to testify, I could and would

24   competently testify to such facts.

25   2.   I understand the Court issued preservation orders in the above-captioned cases,

26   requiring that Google preserve data in ▮▮▮▮ Ads logs, among other data sources. I understand

27   the Court has ordered a daily "sample" preservation of data for 10,000 U.S.-based users in ▮

28   separate ▮▮▮▮ Ads log sources for the *Brown* and *Calhoun* cases, and a "full ongoing

1   preservation" of specific fields from ███████ Ads log sources for the *Calhoun* case. I

2   understand that the Court has also ordered Google to establish a process to decrypt data that is

3   encrypted using the ████████████ re-encrypt the data in a safe and secured manner using

4   a new key created exclusively for this purpose, and preserve the new encryption key for the duration

5   of the litigation.

6   3.   I further understand that Google was required to design and implement the data

7   sampling pipeline for ████ Ads logs by August 15, 2022, to design and implement the

8   decryption/re-encryption pipeline by August 29, 2022, and to complete the backfill of sampled log

9   data going back to July 27, 2016 by September 28, 2022.

10   4.   We have successfully designed and implemented the aforementioned preservation

11   and decryption/re-encryption pipelines and completed the backfill of historical ████ Ads data.

12   To comply with the preservation orders, I, along with two other engineers on the Ad Manager

13   Reporting & Insights Team spent more than ███ hours designing, scoping, configuring,

14   implementing, testing, optimizing, debugging, and verifying data preservation and backfill

15   pipelines. Going forward, those pipelines will require continuous monitoring, maintenance, and

16   debugging, which I estimate will take at least ███ hours per week.

17   5.   The preservation tasks have significantly disrupted my and the other two engineers'

18   regular job duties, which impact Google's ability to maintain existing systems and develop new

19   products in a timely manner. For the past several months, I worked almost exclusively on designing

20   and implementing pipelines necessary to meet the Court's orders, and was able to devote no more

21   than ███ hours per week to my team's regular job duties.

22   6.   The work required to implement the Court's preservation orders requires not only

23   general knowledge of Google's systems and tools, but also specialized knowledge of pipelines and

24   tools to digest historical data and the logs at issue in these cases.

25   7.   To implement the required backfilling of data and preserve historical data for both

26   the *Brown* and *Calhoun* matters in ████ Ads log sources, the preservation pipelines that we

27   designed had to read and process tens of exabytes of historical data to extract the sampled and field-

28   based preservation data. ██████████████████

1   ██

2   ██

3   ██

4   ██

5   ██

6   ██

7   ██

8   ████████████████████████████████ This extreme cost and the

9   bandwidth constraints forced us to find new and nonstandard solutions for allocating resources while

10   minimizing the cost to the company. The amount of Google resources used to implement the court-

11   ordered preservation for ████ logs was █████████████████████████████████

12   ██████, which is an estimate of the RAM necessary to process backfill searches.

13   8.   As of October 13, 2022, the aforementioned sampling preservation pipelines store

14   about ████████████████████ of data, and the compressed daily average of new data

15   written to the pipelines is around ████████ for *Calhoun*; the field-based preservation pipelines

16   store about ███████████████████████████, and the compressed daily average of new

17   data written to the pipelines is around ████████████████ for *Calhoun*; and the sampling

18   preservation pipelines store about ████████████████████ of data, and the compressed daily

19   average of new data written to the pipelines is around ██████████ for *Brown*.

20

21   I declare under penalty of perjury that the foregoing is true and correct.

22   Executed on the 14th day of October 2022 at New York, New York.

23

24   By: *Benjamin Kornacki*

25   Benjamin Kornacki

26

27

28